Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman

Kaj Mäkelä, Esa-Pekka Salonen, Markku Turunen, Jaakko Hakulinen, and Roope Raisamo
Computer-Human Interaction Unit
Department of Computer and Information Sciences
FIN-33014 University of Tampere, Finland
+358 3 215 8558
{kaj, eps, mturunen, jh, rr}@cs.uta.fi

ABSTRACT
Ubiquitous computing is extending the use of information technology to everyday living and working environments. A problem in developing and testing these systems is the fact that they are environments. A complete environment cannot just be taken in a laboratory and tested in conventional usability tests. Here we have addressed the problem of testing such an environment by applying the Wizard of Oz method. This paper describes a Wizard of Oz experiment conducted on a ubiquitous computing system Doorman that is used to control the access of incoming visitors and staff members to the premises and to guide the visitors to find the people or the room they are seeking. The experiment was conducted by simulating speech recognition with a human wizard operating the otherwise fully working system. The user-initiative dialogue strategy was mostly successful, but did not meet the requirements in some cases, as a part of the users were not served properly. The experiment proved to be very valuable in the iterative development of the system.

Categories and Subject Descriptors
B.4.2 **[Input/Output and Data Communications]**: Input/output devices – *voice*.
H.5.1 **[Information Interfaces and Presentation]**: Multimedia Information Systems – *audio input and output*, *evaluation/ methodology*.
H.5.2 **[Information Interfaces and Presentation]**: User Interfaces – *evaluation/methodology*, *interaction styles*.
H.5.3 **[Information Interfaces and Presentation]**: Group and Organization Interfaces – *evaluation/methodology*, *synchronous interaction*.

General Terms
Experimentation, Human Factors

Keywords
Wizard of Oz, ubiquitous computing, spoken language dialogue, speech user interfaces, evaluation

1. INTRODUCTION
The ubiquitous computing applications have extended the traditional desktop paradigm of computing. Daily, people make use of not one but several different computers distributed in their everyday environment. The equipment and applications can be mobile or placed in the environment. More and more ubiquitous computing applications are used in everyday situations.

The use of the ubiquitous computing services is not restricted by the limitations of desktop computing and the context of use can be more unpredictable and informal. One of the main characteristics of a ubiquitous computing application is location awareness. The context of use is formed by a certain location and situation. One way to make this possible is that the system has the initiative and it is able to recognize the needs of the user from the context of the action [13]. Even if location awareness is not used in the system, the context of use and the environment are always important parts of ubiquitous applications.

Because of the nature of the ubiquitous computing, the evaluation of the ubiquitous computing systems cannot be done in a normal laboratory environment. Testing has to be done in the actual scene of action with real life problems. This makes testing the ubiquitous computing environments especially challenging.

The speech-based ubiquitous computing system called Doorman [5] ('Ovimies' in Finnish) is located and being developed in the premises of TAUCHI, the Computer-Human Interaction Unit of the Department of Computer and Information Sciences in the University of Tampere. The Doorman opens the door to the visitors and the staff members and guides the visitors in the premises of TAUCHI.

This paper describes a Wizard of Oz (WOz) experiment conducted during the implementation phase of Doorman system. Wizard of Oz tests are useful in supporting design process and evaluating the interface [2, 3, 14]. The method has been commonly used to test natural language dialogue systems [4] and multimodal systems [8, 14]. Here we apply this method to ubiquitous computing applications.

The paper is organized as follows. First, the description of the Wizard of Oz testing method is provided. The following sections will give information on the Doorman system and the setting of the test conducted for the system. Then, the results of the test are described. Lastly, we discuss the lessons learnt and give suggestions for further research.

2. WIZARD OF OZ METHOD

The Wizard of Oz testing [4, 2] is an experimental user interface evaluation method in which the user of the system is made to believe that he or she is interacting with a fully implemented system though the whole or a part of the interaction of the system is controlled by a human, a wizard, or several of them. The interaction is logged and/or recorded for further analysis.

The Wizard of Oz testing is used to evaluate interaction design and natural language models before they are actually implemented, or can be implemented at the required level of fluency. The testing can therefore support and speed up the iterative development process by directing the development in the right direction.

The Wizard of Oz testing has been found to be suitable for relatively narrow and well-defined application domains in which the application is performing behaviour that can be performed by a human within the available time [2]. Many tasks are faster to carry out for a human than for a computer, but there are also tasks in which the raw processing power of computers is more efficient.

Human-computer communication has been found to differ from human-human communication. Specifically, Baber and Stammers [1] found that humans tend to be polite to each other, but once they know they are discussing with a computer they drop out all the compliments. That is one reason for the fact that all the findings from human-human communication research cannot be directly applied to the human-computer communication. Therefore, to gather reliable information about human-computer communication it is important to observe the human behaviour in a situation in which they believe to be interacting with a real computer system. It is important that the user thinks he or she is communicating with the system, not a human, as noted by Dahlbäck *et al*. [4].

3. DESCRIPTION OF THE SYSTEM

3.1 Overview

The Doorman is used to help the members of TAUCHI and their visitors in their communicational tasks and everyday lives. This is done by automatically:

- opening the door to identified staff members and to visitors, if the target of their visit is recognised,
- guiding visitors in TAUCHI premises to the person or the room they are seeking, and
- giving the staff members information about organisational messages, e-mail, instant messages, phone calls and their visitors who have been asking them when they were absent.

The Doorman uses spoken language to communicate with the users. Speech recognition is used as an input and speech synthesis as an output method. The target of the visitors visit is recognised from their speech using speech recognition. The staff members are identified by recognising their name in their initial phrase. At the moment, speech recognition is the only method for identifying the staff members. The Doorman system has some resemblance to Office Monitor by Nicole Yankelovich and Cynthia McLain [15].

As a ubiquitous computing system Doorman bridges the digital and physical worlds. The system gathers information about the situation at the door with a microphone, a doorbell switch and a door micro-switch. The output of the system is presented to the user with synthesised speech via speakers installed at the door and lobby. The online mode of the system is indicated with a led light next to the doorbell, so that the users know when the system is in use. When the system is offline, the doorbell works in the normal way.

In the lobby the system works in a multimodal way. It uses pointing gestures together with synthesised speech output when guiding the visitor to the target of the visit. The guidance is given by using an anthropomorphic robot pointing to the direction the user should go to find the target. Guidance is formed dynamically and spoken to the user using speech synthesis. The system uses a two-dimensional model of the premises when giving guidance to the visitors. The basic setup of the system is shown in Figure 1.

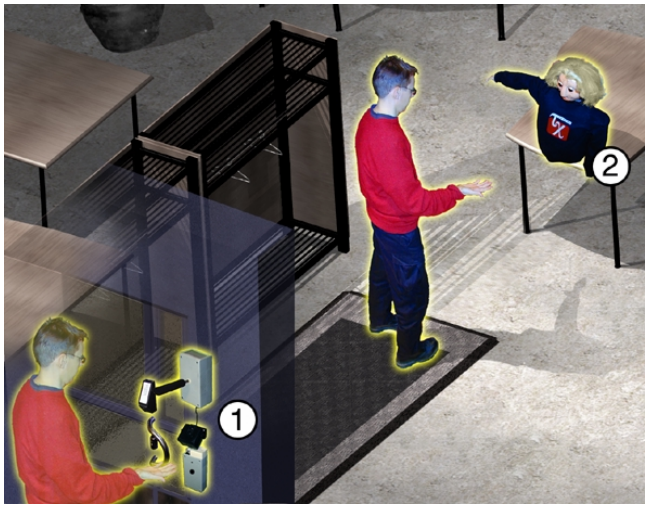

Figure 1. The basic setup of the Doorman system. Outside the door is a microphone, a loudspeaker, a doorbell button and a led light (1). In the lobby there is a guide robot containing a loudspeaker (2).

The Doorman system is based on a distributed software architecture called Jaspis [10, 11]. Jaspis is a Java-based adaptive speech user interface architecture that has been developed in TAUCHI. It was originally designed for spoken dialogue applications, but has been expanded to include features that support developing ubiquitous computing applications.

Jaspis easily enables the Wizard of Oz testing because of its modular manager, agent and evaluator based structure. Each of the components of the Jaspis architecture can be replaced with a wizard. Therefore, implementing the Wizard of Oz version of the system did not require extensive effort. We also added several new features in the Jaspis architecture to support WOz experiments, such as data logging tools. The internal structure of Jaspis is not relevant to the present paper in which we analyse the dialogue of the system. Technical details can be found in the earlier papers [10, 11].

3.2 Dialogue model

The dialogue control model of the system is implemented as a finite state machine. Each system state (usually one turn in a conversation) is implemented as an independent dialogue agent. Any modifications were not needed because of the WOz experiment in the dialogue agents or presentation agents producing the speech outputs. Therefore, the system was fully functional except for the fact that a wizard simulated the speech inputs. The wizard was not able to control the behaviour of the system in other ways.

The functions of the current system can be divided into following stages:

1. a) Recognition of the staff member, or
   b) Recognition of the target of the visit.
2. Opening the door, and
3. a) Greeting the staff member, or
   b) Guiding the visitor to the target of the visit; the target can be a person or a room.

In the current dialogue model the system prompts are formed to guide the user to answer briefly. The visitors are assumed to push the doorbell, after which they are asked and expected to tell the target of their visit. It is assumed that the staff members will not push the doorbell but say a greeting and their name straight away on the door. The structure of the dialogue is presented below. The dialogue is translated from its original form that is spoken in Finnish.

A. Staff Members

1a. if the user speaks, go to step 2
1b. if the doorbell button is pushed,
DOORMAN: "What is the name of the person or the room you are searching?" // Target request
2. STAFF: "John Doe here, hello. Could you open the door?" // Name and greeting
3a. If the name is recognised, // confirm
DOORMAN: "Good morning, John Doe. I will open the door for you."
open the door
3b. if the name is not recognised, go to step C1
4. DOORMAN: (inside) "Good morning, John Doe."

B. Visitors

1a. if the user speaks, go to step 2
1b. if the doorbell button is pushed,
DOORMAN: "What is the name of the person or the room you are searching?" // Target request
2. VISITOR: "John Doe" or "usability lab" // Name of the target
3a. if the name is recognised,
DOORMAN: "Good morning. I will open the door for you."
The door is opened.
3b. if the name is not recognized, go to C 1.
4a. if the target is a person,
DOORMAN: "Good morning. The person you are searching is in room 444. To find there turn left, go seven meters forward, turn right, go five meters forward, turn left. The room of N.N. that you were looking for is on right seven meters from you."
4b. if the target is a room,
DOORMAN: "Good morning. To find your way to the usability lab, which is the room number 412 turn right, go three meters forward, turn right. The usability lab is nine meters ahead. "

C. Error handling

1. DOORMAN: "I am sorry, I did not understand. Say the name of the person or the room you are searching for."
2. VISITOR: --- // statement that cannot be recognised
Consecutive errors are responded in the following way:
3. DOORMAN: "Say the name of the person or the room you are searching for."
Step 3 is repeated three times.
4. DOORMAN: "I am sorry, I cannot open the door. Use the key or push the doorbell button within 15 seconds to ring the doorbell."

These scenarios demonstrate all the cases that the tested system was able to handle. In case the users would try out a different strategy, error correction in scenario C is intended to handle the situation. The system uses user-initiative dialogue control strategy by default and takes the initiative when the user pushes the doorbell.

4. DESCRIPTION OF THE EXPERIMENT

4.1 Goals

The aim of the study was to test and analyse the spoken language and multimodal dialogue model designed in the system before constructing the actual speech recognisers.

The study consisted of two parts:

1) The use of speech synthesis and spoken language in ubiquitous computing application. The interesting point was to find out what kind of language users actually use when talking to this kind of computer system. It was also interesting to find out how the form of the speech output affects on the behaviour of the users and the language they use to communicate with the system. This is useful information for the design process of the vocabulary, the grammar of the dialogue and the interaction model.

2) Combining synthesised speech and pointing gestures in guiding the users to different rooms. The interesting point was to find out how the route to the target of the visit should be given to the user so that the user would comprehend the guidance. This contains prosody and timing of the speech and synchronisation of the movements and instructions.

The aim of the test was to recognise the actual behaviour of the user and the problems occurring in the following situations:

- the user understanding the question given by the system using synthesised speech,
- the visitor responding to the question and stating the target of the visit,

- the staff member declaring his/her identity,
- the behaviour of the user while entering the premises and
- the visitor understanding and responding to the guidance given by the system.

The collected data was to be used to evaluate how well the current dialogue model works, and to give insight to how to improve it when the system is further developed.

4.2 Experimental setup

The test was conducted in five days, one of which was used for training and pilot testing the setup. The test was run approximately 4 hours per day, on a quite varying basis. The test sessions lasted from 45 minutes to 1.5 hours each time. The test was conducted by two persons: one was acting as a wizard and one was gathering permissions from visitors for recording. The test group members changed roles many times during the test, because the wizard task was quite demanding and their alertness would go down in a long-lasting session.

The visitors were informed of the system with posters next to the system outside the door. In the poster it was explained that the system uses speech recognition and that it can be bypassed by pushing the doorbell button three times in a rapid sequence.

The staff members were informed of the system via e-mail before the experiment was started. However, the nature of the system was not revealed. They were asked a permission to gather voice samples and log information to create their personal profile. These voice samples will later be used to improve speech recognition accuracy and to recognize the users from their voice.

The staff members were not given any specific instructions on using the system. They were only told to greet and introduce themselves to the system to see which way they would behave without detailed instructions.

The Doorman system opens the door for the users. The door can also be opened with traditional keys or with electronic key cards. Key owners were asked but not forced to use the system. This request may have had an effect on the naturalness of the test setting. However, this was necessary to ensure gaining data on the use of the system.

According to the Finnish law it is required to inform the users when their speech and actions are recorded. The permissions of the users were gathered in written form. Because of this, one person of the test group was in the lobby gathering permissions from visitors after the entrance. The permissions of the staff were gathered before starting the experiment.

The users were not told that the system was controlled by a human wizard, because the information could have been spread and affected the results. The staff members were told about the Wizard of Oz testing after the experiment.

The wizard listened to the audio input gathered with a microphone, synthesised speech generated by the system and the doorbell signal. The same audio information was recorded for analysis. The testing equipment was situated and operated in a separate room with no visual contact to the door or the lobby. Figure 2 shows the high-level structure of the setup used in the WOz experiment.

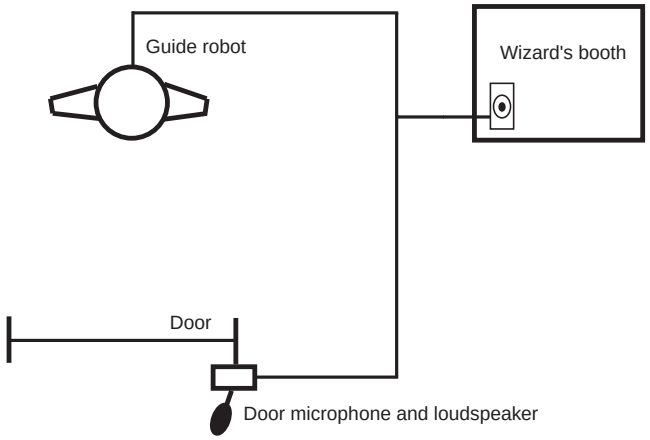

Figure 2. The experimental setup.

5. WIZARD TOOL USER INTERFACE AND WIZARD RULES

The Wizard of Oz experiment was conducted by replacing the forthcoming speech recognition module of the Doorman system with a control application used manually by the wizard observing the situation. The speech of the users was recorded and all the system tasks and sensor inputs were logged to be thoroughly analysed later.

The wizard listened constantly to the voice input gathered with the microphone installed to the door. When a user spoke the wizard interpreted and conveyed the input to the system.

5.1 User interface of the control application

We implemented a tool for the wizard to give speech recognition information manually to the system. The interface of the tool is shown in Figure 3. The tool is similar to the wizard tools used in previous experiments, as by Dahlbäck *et al*. [4]. The control application was designed to be as simple to use as possible to ensure short response times and to minimise the possibility for errors. The wizard control application was implemented as a Java applet. In relation to the rest of the Jaspis-based system, it acted as a speech recognizer connected by using socket connections.

The tool provides a simple list-based user interface consisting of all the possible alternatives for the speech recognition results. The lists are fitted in one screen so that there is no need for scrolling. The tool converts the chosen option to the string form and sends it to the system. Basically, there are three kinds of inputs: the identity of the staff member, the target of the visitor's visit, and the recognition errors.

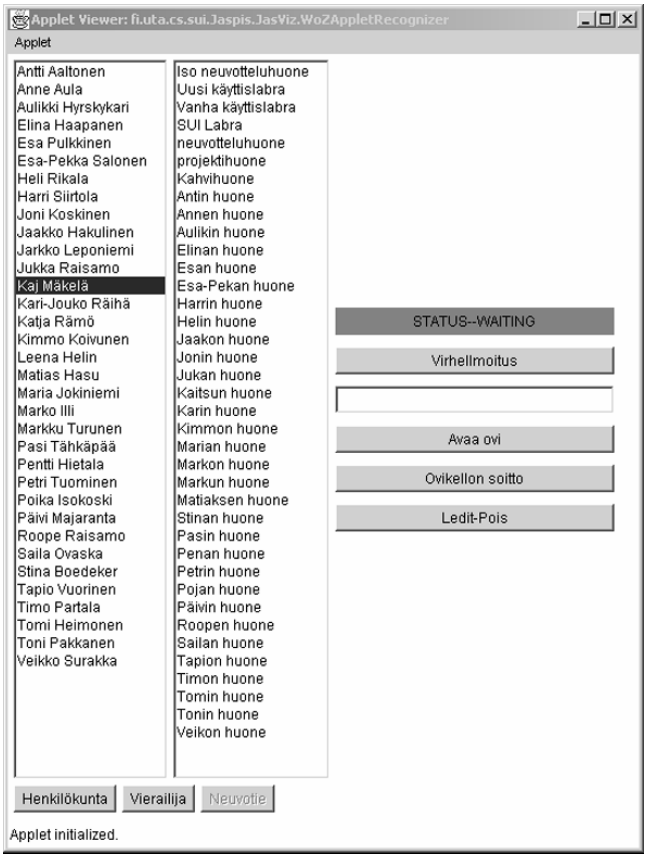

Figure 3. The user interface of the control application. The two lists contain all the staff members (on the left) and the rooms (on the right) in TAUCHI. They are used to identify the user as a staff member ("Henkilökunta") or to select the target of the visit ("Vierailija" or "Neuvo tie"). There is also a button for error messages ("Virheilmoitus"). For exceptional cases there is a possibility to open the door manually ("Avaa ovi") or call for help by ringing the doorbell ("Ovikellon soitto").

The interface in Figure 3 contains two single-selection lists, seven buttons and one indicator. The user's identity or the target of the visit is chosen from the lists. There is one list consisting of the names of the staff members and another consisting of the room names. When the name of the person is chosen from the list a different button is used to tell the system whether the chosen name is a target of a visit or an identity of a staff member. When the target of the visit is a room, the name of a room is chosen from the list and submitted with a button. There is also a button that sends the system a recognition error. This is used when the speech of the user does not contain the information expected. For unpredictable situations the wizard can use a button to ring the doorbell or a button to open the door.

There is a status indicator providing the wizard with information about the mode of the system. The indicator is red if the system is waiting for input from the wizard and green if the system is handling the given input. The wizard is also given the information of the state of the door and the actual doorbell button via the feedback sounds played at the door. The wizard system can be enabled and disabled using the tool by pushing a toggle button.

It should be noted that the system was otherwise fully functional and the wizard was unable to alter the behaviour of the system in any other means except giving simulated speech inputs. Furthermore, the simulated speech inputs were always either legal inputs or indicated recognition errors. Only one kinds of recognition errors (not recognized) were simulated mainly because we wanted to simplify the WOz experiment and the work of wizards.

5.2 Wizard rules

To keep the behaviour of the system consistent and credible we formed a set of rules for the wizards operating the system. One of the main problems in the Wizard of Oz testing is that the wizard has superior knowledge and skills compared with the system being simulated, and he or she has to reduce skills and knowledge to emulate a software component [2]. The following rules aim at resolving this problem in our experiment.

1. The speech of the user should contain any name of a staff member or room in TAUCHI premises, otherwise it will cause a recognition error. However, the common sentence structure should be used.
2. There should be only one person speaking at a time.
3. The staff members are required to say something more than just their own name. The speaker is recognised as a staff member if the utterance contains, for example, a greeting of some kind or a request to open the door. This rule was formed because the visitors express the person they are searching by saying only the name of the person. Therefore there was a need to differentiate the ways to identify a person since the system does not currently have other speaker recognition capabilities. It was decided that the person saying the name alone would be identified as a visitor.
4. The Wizard should react to everything that is said at the door, even to a speech that is not necessarily targeted to the system. This is to give expression of continuous speech recognition.

These rules were carefully followed during the whole experiment.

6. FINDINGS

We made several observations during the experiment concerning the behaviour of the wizard and the users.

6.1 Behaviour of the Wizards

Two persons acting as wizards managed to keep the operation of the system consistent and correct. The biggest problem was to rapidly decide during the experiment how to handle unexpected speech inputs. The wizard rules were updated and discussed between the wizards whenever new problems had been found. Part of the rules mentioned in Section 5.2 was formed during the experiment. This was the case, for example, on rule four, which was formed during the practice.

The tool itself was simple enough to enable quick and consistent responses. Only one input mistake occurred during the testing. The pace of responses of the wizards was kept the same. However, the system delays caused small variation in response

time of the system. The content and style of simulated inputs was predefined and did not vary during the test.

6.2 Behaviour of the users

The testing sessions were recorded and analysed afterwards. There were three distinct groups of users: 1) the visitors, 2) the students and staff members who do not belong to TAUCHI, and 3) the TAUCHI staff members. During the experiment, the system was used in 74 occasions, of which 22 were visitors, and 52 were staff members. It was not possible to distinct the students and other university staff members from the visitors using the system. Therefore they are also handled as visitors in these statistics. However, they have clearly different needs and usage patterns and because of this we handle them as separate groups whenever possible.

Fifteen visitors (68 percent) used the system so that they responded to the first prompt in the way they were expected to. One user succeeded in the second try and one in the third try. Three users bypassed the system by pushing the doorbell three times. Two users were not able to get in by using the system. The result shows that the system prompt was formed so that in most of the visitor cases (77 percents, 17 persons) the users answered in the way they were expected, and thus the system successively served these visitors.

The visitors were assumed to come to TAUCHI to meet someone or to find some room in TAUCHI premises, for example the meeting room or the usability laboratory. In the system prompt the users were informed that they could state a name of a person or a room. However, during the test all the visitors were searching for a person, not a room. For example, a visitor coming to a meeting held in a main meeting room stated that he or she is coming to meet the staff member organising the meeting.

The visitors did not have a key or a key card to the premises and therefore they normally used the system in the visitor mode that was triggered by the doorbell button. The visitors were given a possibility to bypass the system and ring the doorbell by pushing the doorbell button rapidly three times in sequence. There were four situations (of total 22, 18 percent) in which the visitor did not want to use the system and three of them (14 percents) used this possibility.

Students and the university staff members who do not belong to TAUCHI have a key card they can use to access TAUCHI premises. Therefore, they had a possibility to bypass the system and they mostly used this possibility despite our written request to use the Doorman system. This may have been because of the routine when they had visited many times earlier before the system was available, or because of the ease and the quickness of use of the key card. The students and these staff members were not necessarily even aware of the existence of the system, because the printed poster may have been unclear or too long to be read. The students and university staff members also know the premises and need no assistance in finding a person or a room. So when they did not use the system it only implies that they did not need it, whereas the visitors had a real need for the system.

We found out that the staff can be divided into three groups from the system use point of view. One group is the users using the system regularly, most of the times when coming in. We call them active users. They were mostly people visiting outside the premises often, for example to smoke. It is also customary to help the fellow researches in their research by voluntarily assisting them in gathering data. This may have had an effect on the use patterns of the active users.

TAUCHI is an expert organisation, where all the employees are experts of some area of usability and interaction. This was also shown in the behaviour of the active users. Some active users were constantly testing the abilities of the speech recognition engine by using complex impressions and several users speaking at the same time. However, it was shown that these users learned the restrictions of the system and learned to use the system.

The second group consisted of those who tried the system only once or few times. They were interested to know how the system works, but lost their interest quite soon.

The third group was those who did not use the system at all, but instead used other methods to get inside the premises. There are several reasons for some people not using the system. There is the possibility to use keys and key cards to enter the premises. The members of the staff are accustomed to use keys and key cards and often used them. There is also another entrance to the premises, which is used by those staff members who have their office on the other side of the premises. The speech output of the system took quite much time (mean 6 seconds) and there were also some small delays in the system. The whole process from the start of the speech of the user to opening the door took approximately 16 seconds (mean). This includes the speech of the user, the system and wizard delays and system prompts. The users were also informed that the system is gathering data of their behavior and they may have been avoiding the system because of this.

In the dialogue model it was assumed that the staff members are not willing to push the doorbell button in order to get inside. This is why they were given a possibility to introduce themselves at any time the system was in standby mode. However, in 19 percent of the cases the staff members did use the doorbell and therefore heard the prompt formed for the needs of the visitors. Some of the staff members used this manner multiple times and did not recognise the problem.

6.3 Discussion of the Findings

The system was able to serve 68 percent of the visitors at the first try and 77 percent after the second or third try. These users acted as had been expected. The result is promising for the design process, as this was the first iteration from the usability point of view.

The aim of the system is to serve all the users in some way, at least by calling for external help when a problem arises. However, during the tests the system failed to serve the user in two cases. Although this is a small amount, it should be seriously considered. In one case, a user did not speak to the system at all and did not use the possibility to bypass the system. In the other case, the person the visitor was searching for was not a staff member. In this case, the user got frustrated when his speech was not recognized, and used his cellular phone to contact the person he was supposed to meet. In this case, the user tried three times and stopped after this. The system was programmed so that after the fourth try it will go to a state where the user can push the doorbell button to actually ring the bell inside. This observation leads us to the assumption that the system should go to the manual mode after three or already after two failed recognitions and ring the bell in a

normal way. Furthermore, we should have other ways to handle situations of this kind.

The other problem in this case was that the system is able to guide the visitor only to the members of the TAUCHI staff listed in the system. However, the usability laboratory is also used by the students and other organisations. The testee arriving to the test will then be searching for the person who is conducting the usability test and the system is unable to recognise the person searched. It is also possible that the person coming to the test does not even know the name of the person conducting the test or the name of the room the test is held in. It is very difficult to detect automatically when the users state names not known to the system. This out-of-vocabulary detection is one of the weak points of the current speech recognition systems. It is also out of the question to list every possible option to the user using spoken prompts.

During the experiment, the visitors did not use names of the rooms when stating the target of their visit. However, the experimental data is quite limited and further testing should be conducted before drawing generalising conclusions.

The delays in the system response were found irritating and the users having a key or a key card often chose to use one instead of waiting the system to react. This was partly because the user was not sure when the system was processing the input due to lack of any indicator or feedback showing the current state of the process to the user. The human wizard, detailed event logging and limited hardware resources caused some delays. Also the length of the sentences spoken by the system annoyed the users. Especially the users using the system on a regular basis were irritated, because the speech delayed their entrance.

The guide robot was often passed without listening to the instructions. The reason mostly was that the person already knew where he or she was going. The other reason was that the robot and the guidance were not given consideration. The robot had also been on location long before the system was functioning and the users may have not expected it to act. The timing caused that the robot started guiding too late and the visitor had already passed it.

The guidance given by the guide robot was also found too long, slow and unclear. The timing, the length of the prompt and the speech rate altogether caused that most of the visitors ignored the guidance. The length of the guidance also made it hard to remember the guided route.

Some of the staff members started the dialogue by greeting the system and waited the system to respond before stating their name. This may have been because they wanted to make sure that they are heard and to make sure that the connection with the system is established. In these cases, the users behaved much like in human-human interaction and expected the system to behave similarly. They expected the system to be able to work in a more sophisticated level of conversation than it really was. This is one of the learned communication patterns which people use in their daily communication. Even if people know the limits of the system they use their learnt skills. We also gave instructions for staff members to greet the system.

It was shown that the people using the system on more a regular basis changed their way of speaking to the system by their former experiences. They learned from their mistakes and adapted their interaction to the system. This is consistent with the observation made by Tennant [9]. However, since some people stopped the use of the system after few attempts, it is possible that they do not want to adapt to the system.

In informal conversations the users told that the speech synthesis was unclear and therefore sometimes hard to understand. It has been shown that listening to synthesized speech requires more processing capacity than listening to natural speech before human has encoded the synthesized speech. Therefore, we assume that the people, who had difficulties in understanding synthesized speech, simply were not accustomed to hear it. Our observation is consistent with remarks by Weinschenk and Barker [12, pp. 190-191].

It was shown that the users will choose the easiest and quickest way to handle the task and if the system is not able to serve them properly they will choose an alternative method.

7. LESSONS LEARNT FROM THE EXPERIMENT

As a part of the iterative design process, some improvements are going to be made to the Doorman system on the basis of the test. To make the interaction with the system quicker, and to have more people to use the system, the level of the system initiative is going to be increased by implementing sensors to recognise the presence of the user on the door. This is used to trigger the dialogue.

The results show that some additional consideration should be given in the form of the speech output and the delays of the system. Long and predefined system utterances are going to be shortened, varied and made more informal and natural, for example, by utilising the user profiles of the staff members and varying greetings. The speech rate will be increased to make the speech more understandable and shorter. This also makes the speech shorter. In addition, the intonation of the speech can be altered to make the speech clearer.

The order of speech and action is going to be rearranged to shorten the delays, and the flexibility of dialogue is going to be increased by giving the user a possibility to interrupt the synthetic speech. The user should be given feedback after receiving the speech input to confirm that the system is reacting. This is going to be done by using short utterances like, for example, ‘hmm’ or ‘let’s see’.

In some problematic situations, when the user did not use correct words, the interaction failed. This happened, for example, when he or she said something beyond the vocabulary of the system. The error loop after a failed speech recognition attempt was found to be too long and it is going to be shortened. Each turn of the loop should also adapt to the situation and give more detailed instructions to the user. In the dead-end situation, the system should be able to offer alternative solutions such as to call human operators.

We should slightly alter the dialogue model to better support system-initiative dialogues. For example, many staff members did not take initiative by speaking, but instead acted like visitors and pushed the doorbell button. It also seems that visitors need more guidance and system-initiative dialogue.

The guide robot and the guidance messages were unsuccessful and we should really focus on these issues. The guidance messages were found to be too long and too confusing. Especially we need

to change the guidance messages to use landmarks instead of direct walking instructions. The appearance and position of the guide robot should be altered to make it more noticeable. The guide robot will be more visible to make it better noticed. The association between the speech outside the door and the guide robot needs to be evident.

The test also brought up a need for changes in the architecture level: at the moment, the Jaspis architecture does not support simultaneous dialogues. The changes that allow this feature are going to be implemented in the near future. Also, the system delays are going to be shortened by optimising the system functionality and the architecture.

To gain more results on the use of the system, the amount and activity of the users should be increased. This could be accomplished by running the test day and night for an extensive period of time. However, this would require wizards controlling the process all the time and would be a highly laborious task. This suggests that carrying out one or more of this kind of limited-time Wizard of Oz experiments could make better use of the resources during the development process of the system than one extensive study. Instead, more extensive study would be more valuable from the interaction analysis point of view.

In future we are going to arrange more tests, some of them for visually impaired users to find out how they use the system and especially how the guidance system works with them. We are also going to implement more WOz support features into the Jaspis architecture on the basis of this experiment.

8. CONCLUSIONS

This paper described a Wizard of Oz experiment that was used to find out the way the users interact with the Doorman system. Even informal, the results are useful in giving guidelines to the following iterations of the design process.

The most important findings that help in the further development of the system were related to the structure of the dialogue, the need for system initiative and better error handling, and the way the guidance is arranged. The guide robot needs to speak in common terms and to be easily recognizable.

The experiment gave us valuable information on how to improve the system. It also showed that setting up a Wizard of Oz experiment did not require extensive modifications in the tools provided by the Jaspis framework. Based on our experience we recommend using Wizard of Oz method during the iterative development of ubiquitous computing systems.

9. ACKNOWLEDGMENTS

The Doorman system was implemented by Jaakko Hakulinen, Anssi Kainulainen, Kaj Mäkelä, Esa-Pekka Salonen and Markku Turunen. The project was lead by Professor Kari-Jouko Räihä. We would like to thank Leena Helin for assisting in the WOz experiment and Saila Ovaska for her comments during the writing process. This work was carried out in the project 'User Interfaces for Ubiquitous Computing' funded by the Academy of Finland (project 163356).

10. REFERENCES

[1] Baber, C. and Stammers, R.B., Is it natural to talk to computers: an experiment using the Wizard of Oz technique. In E.D. Megaw, (Ed.), *Contemporary Ergonomics 1989*. Taylor & Francis, 1989.

[2] Bernsen, N.O., Dybkjær, H., and Dybkjær, L., *Designing Interactive Speech Systems*. Springer-Verlag, London. 1999, 127-160.

[3] Bretan, I., Ereback, A.-L., MacDermid, C., and Waern, A., Simulation-based dialogue design for speech-controlled telephone services. *CHI'95 Interactive Posters*, available at http://www.acm.org/sigs/sigchi/chi95/proceedings/intpost/ib_bdy.htm.

[4] Dahlbäck, N., Jönsson, A., and Ahrenberg, L., Wizard of Oz Studies – Why and How. *Proceedings of the 1993 International Workshop on Intelligent User Interfaces (IUI'93)*, ACM Press, 1993, 193-200.

[5] The Doorman system http://www.cs.uta.fi/hci/spi/Ovimies/index_en.html.

[6] Gustafson, J., Bell, L., Beskow, J., Boye, J., Carlson, R., Edlund, J., Granstrom, B., House, D., and Wiren, M., AdApt-A Multimodal Conversational Dialogue System in an Apartment Domain. *CD ROM Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.

[7] Johnsen, M., Svendsen, T., Amble, T., Holter, T., and Harborg, E., TABOR – A Norwegian Spoken Dialogue System for Bus Travel Information. *CD ROM Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.

[8] Salber, D., and Coutaz, J., Applying the Wizard of Oz Technique to the Study of Multimodal Systems. Proceedings of the EWHCI'93, Third International Conference. *Lecture Notes in Computer Science*, Vol. 753, Springer-Verlag, 1993, 219-230.

[9] Tennant, H., *Evaluation of Natural Language Processors*. Ph.D. Thesis, University of Illinois Urbana-Champaign, 1981.

[10] Turunen, M., and Hakulinen, J., Jaspis - A Framework for Multilingual Adaptive Speech Applications. CD ROM *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.

[11] Turunen, M., and Hakulinen, J., Agent-based adaptive interaction and dialogue management architecture for speech applications. In Text, Speech and Dialogue. *Proceedings of the Fourth International Conference TSD 2001* (to appear).

[12] Weinschenk, S., and Barker, D.T., *Designing Effective Speech Interfaces*. John Wiley & Sons. 2000.

[13] Weiser, M., The Computer for the 21st century. *Scientific American*, 265(3). September 1991, 94-104.

[14] Wyard, P., and Churcher, G., A realistic Wizard of Oz simulation of a multimodal spoken language system. *Proceedings of 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, 1998.

[15] Yankelovich, N., and McLain, C.D., Office Monitor. *CHI 1996 Conference Companion*, ACM Press, 1996, 173-174.